Editorial

Accounting for scale and randomness in patterns of forest responses

1. Introduction

The science and practice of forest ecology and management rely on the use of statistical analyses to understand the processes and patterns of forests over time and across landscapes. Powerful, sophisticated approaches to statistical examination of data are necessary for accurate and precise insights. Sometimes a focus on details of techniques can be undermined if insights from a true pattern are misapplied to a different scale, or when important characteristics of sampling and distributions are overlooked. Three case studies in this editorial examine some of these challenges, in hopes of reducing some occasional errors of interpretation in manuscripts. The first case examines how the response of forests across a geographic area may not provide a good representation of the response of individual forests. The second explores how resampling a population can erroneously suggest that large reductions in values occurred for sites that previously had high values, while increases occurred in sites with low values (the null expectation of regression to the mean). The last case shows how statistical designs that test for differences in means may be over interpreted if the distributions are contrasted without adequately considering null expectations.

2. Case 1. Patterns across sites may not apply within sites: growth responses of Eucalyptus to water supplies

The growth of most forests is limited commonly by the supply of water. Forest growth typically increases in wetter years, in response to irrigation, and across geographic gradients of increasing precipitation. How much would growth of a stand change if water supply increased by 100 mm yr^{-1}? A first approximation might be based on the increase in growth for a wet year compared to a normal precipitation year. A *Eucalyptus* plantation in Bahia, Brazil showed a strong increase in stemwood growth for a wet year, increasing by 3.2 Mg ha^{-1} yr^{-1} for a 100-mm increase in water supply from rain (Fig. 1; Stape et al., 2008). It might be tempting to expect that irrigation during a normal (or dry) year would give a similar response, but in fact the response to irrigation during a normal precipitation year was only 2.4 Mg ha^{-1} yr^{-1} for a 100-mm increase in water input. An extrapolation from the precipitation effect between years to the irrigation effect within a normal precipitation year would have overestimated growth response by one-third.

The growth of *Eucalyptus* plantations across geographic gradients in precipitation might be expected to match these within-site responses to water supply, but in fact the growth responses to increasing precipitation across Brazil were only 1.5 Mg ha^{-1} yr^{-1} for each 100-mm increase in precipitation (Binkley et al., 2017, based on 27 sites with 12 clones are each site). The extrapolation from within-site responses, based on either irrigation or rainfall, would overestimate the geographic rainfall effect by 60% to more than 100%.

Growth responses to water supply are moderated by temperature and humidity. Atmospheric vapor pressure deficits may be higher on drier sites than wetter sites, limiting the ability of trees to respond to increases in soil water availability. Geographic locations with low rainfall are often warmer with greater vapor pressure deficits that would likely constrain the ability of trees to respond to increases in soil water availability.

This example could also be turned around. The response across the geographic gradient of precipitation might be used to infer the likely response within a single site. The example in Fig. 1 shows this would be poor estimation of the actual response to precipitation or irrigation for this single site. The prediction of responses at single sites is always more uncertain than the responses of many sites, as the local details (genotype, temperature, humidity, soil, silviculture and other factors) shift the observed value through the broad range of the variance of the overall population. Extrapolation across scales, from single sites to populations and vice versa, may have some value. However, the fundamental importance of confounding factors always warrants explicit consideration. (This issue is an example of a problem that is sometimes called Simpson's Paradox.)

3. Case 2. Resampling may lead to random patterns that look suspiciously like signals: accounting for regression to the mean in tracing changes in soil carbon

A core aspect of statistical analysis is that the value for an observation depends on both a signal and noise. A signal might be the true effect of an experimental treatment. Noise might result from a variety of sources, including variation in properties within a population, or from factors not examined in the experiment. If the ratio of signal to noise is high, we have confidence that an experimental factor influenced the size of the value we observed. When the noise is large relative to the influence we can attribute to an experimental factor, we acknowledge that the experiment does not lend support for the original hypothesis.

These issues of signal and noise are of course at the heart of classical statistics, but in some cases it's surprising that "noise" may appear as an impressive pattern rather than a diffuse cloud of data points. A recent paper in a top science journal examined how heating of forest soils might lead to increases or decreases in the pool of carbon stored in the soils. The authors concluded there was a gain of carbon when heated soils had initially low levels of carbon, but that soils with high carbon initially lost carbon. Should we be confident that soils high in carbon
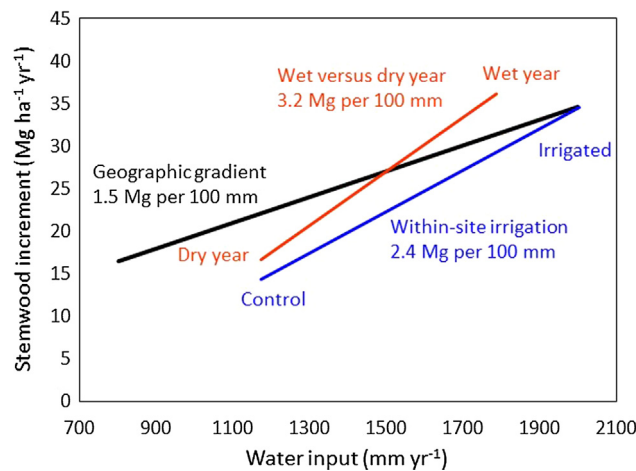
Fig. 1. The growth of Eucalyptus plantations increased by about 1.5 Mg for each increase of 100 mm across a broad geographic gradient in Brazil (Binkley et al., 2017). This pattern did not match the size of the response within a single stand, to either supplemental water from irrigation (2.4 Mg per 100 mm extra water), or to the response to wetter-than-average weather (3.2 Mg per 100 mm, likely because of lower vapor pressure deficits in the wetter year; data from Stape et al., 2008).

will show higher losses of carbon as climates warm? Aside from challenges posed in warming soils in a realistic way, this question needs to be considered carefully for the possible effect of noise. A well-known, commonly overlooked issue in determining changes over time is called "regression to the mean" (see for example, Barnett et al., 2005).

A sampling approach typically defines a population of inference, and samples are then taken from that population (often at random) to determine characteristics such as the mean and the dispersion of observations around the mean (the variance). Statistical tests can contrast sets of sample from two populations (perhaps one a control, and one subjected to an experimental treatment) and give a probability that the samples come from different populations (such as whether a treatment effect was significant). If we ask whether the carbon content of soils from multiple sites changed over time, the two populations being compared come from the same sites but at different times. If classical statistical tests show that we can be confident the soils changed over time, we might then be curious if some soils changed more than others, and ask questions like the one above about whether the change in carbon related to the initial carbon content of a site's soil.

This is the question asked by Cook et al. (2016). They sampled over 300 locations across Brazil for the carbon content of soils (0–30 cm depth), and repeated the sampling after 3 rotations of *Eucalyptus* plantations. Across all sites, the average change in soil C was a decline of $0.22\,Mg\,ha^{-1}\,yr^{-1}$. Did sites with higher C contents show higher losses? A plot of the change in soil C as a function of the initial soil C content (Fig. 2) showed that a site that started with a very low amount of C (such as $10\,Mg\,ha^{-1}$) tended to gain C (about $1\,Mg\,ha^{-1}\,yr^{-1}$). A site that initially had high C (such as $60\,Mg\,C\,ha^{-1}$) tended to lose C (a change of $-1\,Mg\,ha^{-1}\,yr^{-1}$). This trend was very robust, with an r^2 of 0.56; the probability of the relationship was only random (the P value) was less than 10^{-50}.

The value for the change over time for each site included a true signal: the actual change for the site. Each value also included noise, including random variation in how well the sampled soil represented the true value for each site. In the original sampling, a site may have had a high value because of a combination of true soil C content and random error. The resampling would again be influenced by the true value for each site, along with the random noise associated with the sampling. The key point is that a site that truly had a high C value in the first sampling would likely have a high C value in the second sampling, but a site that had a high value originally because of random noise would not repeat with a high value. The signal would remain the same between samplings (plus or minus any real change in the true value), but the noise would be randomly redistributed across the sites giving

lower values for formerly high-value sites, and higher values for formerly low-value sites.

Could this regression to the mean be responsible for the change in soil C observed across the 300 sites? This question was raised by Mats Olsson at the Swedish University of Agricultural Sciences, and Cook et al. (2016) examined the expected pattern of change that might result from only random variations among the 300 sites. They took the initial value for each site and then created a random number for the "re-sampled" value. A column of numbers for the resampled values was generated using the overall mean of the 300 sites, and the standard deviation. The differences between the initial values and the randomly created new values showed the expected pattern. Sites with initially high values in C tended to be matched with random numbers that were closer to the mean (or below it), indicating soil C declined. Sites that were initially low in soil C tended to be matched randomly with values that were closer to the mean (or above it), indicating that soil C increased. A plot of these differences between the real values for the initial soil sampling and the randomly generated "resampled" values was indistinguishable from the pattern shown by the real resampling of the sites. Cook et al. (2016) concluded that a pattern that seemed to have a one in 10^{-50} chance of being random was indeed the random outcome of regression to the mean. Sites with high initial soil C were not in fact likely to have more or less change in soil C than other sites. The vanishingly small P value actually gauged the likelihood that 2 sets of 300 random numbers would correlate with each other (which of course is unlikely), providing a critical indication of the value of considering null expectations for data sets.

4. Case 3. Null expectations need to be considered carefully: limits on interpretation of twin-plot and triplet designs

Regression to the mean deals with paired sampling in time, where a new value is determined for a previously sampled unit. Sampling may also be paired in space, with adjacent (or split) plots assigned as a control or a treatment. This paired plot approach can be quite useful, especially as an inexpensive approach for testing a factor across a landscape at the scale of management programs (Binkley et al., 2018). Fertilization responses have been examined as twin-plots (Stape et al., 2006), and a similar triplet design has been used to see how growth in mixtures of two species compares with the growth of monocultures of each species (Pretzsch et al., 2017).

The key feature of a twin-plot or triplet design is that no replication of treatments is provided within each site; all degrees of freedom are dispersed across the larger population of inference represented by the